

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number

UTILITY PATENT APPLICATION TRANSMITTAL Only for new provisional applications under 37 C.F.R. 1.53(b)	Attorney Docket No.	NTL-3.2.113/3199 (11334RO)
	First Inventor or Application Identifier	Elias J. Nemer
	Title	METHOD OF AND APPARATUS FOR REDUCING ACOUSTIC NOISE IN WIRELESS AND LANDLINE BASE TELEPHONY
	Express Mail Label No.	EL284833165US

APPLICATION ELEMENTS See MPEP chapter 600 concerning utility patent application contents.	ADDRESS TO: Assistant Commissioner for Patents Box Patent Application Washington, D.C. 20231
---	---

<p>1. <input checked="" type="checkbox"/> *Fee Transmittal Form (e.g., PTO/SB/17) (submit an original, and a duplicate for fee processing)</p> <p>2. <input checked="" type="checkbox"/> Specification Total Pages [37] (preferred arrangement set forth below)</p> <ul style="list-style-type: none">- Descriptive title of the invention- Cross References to Related Applications- Statement Regarding Fed sponsored R&D- Reference to Microfiche Appendix- Background of the invention- Brief Summary of the invention- Brief Description of the Drawings (if filed)- Detailed Description- Claim(s)- Abstract of the Disclosure <p>3. <input checked="" type="checkbox"/> Drawing(s) (35 U.S.C. 113) Total Pages [6]</p> <p>4. Oath or Declaration Total Pages [1]</p> <p>a. <input checked="" type="checkbox"/> Un-executed (original or copy)</p> <p>b. <input type="checkbox"/> Copy from a prior application (37 CFR §1.63(d) (for continuation/divisional with box 16 completed)</p> <p>i. <input type="checkbox"/> <u>DELETION OF INVENTOR(S)</u> Signed statement attached deleting inventor(s) named in the prior application, see 37 CFR §§ 1.63(d)(2) and 1.33(b).</p> <p>NOTE FOR ITEMS 1 & 13: IN ORDER TO BE ENTITLED TO PAY SMALL ENTITY FEES, A SMALL ENTITY STATEMENT IS REQUIRED (37 C.F.R. § 1.27), EXCEPT IF ONE FILED IN A PRIOR APPLICATION IS RELIED UPON (37 C.F.R. § 1.26).</p>	<p>5. <input type="checkbox"/> Microfiche Computer Program (Appendix)</p> <p>6. Nucleotide and/or Amino Acid Sequence Submission (if applicable, all necessary)</p> <p>a. <input type="checkbox"/> Computer Readable Copy</p> <p>b. <input type="checkbox"/> Paper Copy (identical to computer copy)</p> <p>c. <input type="checkbox"/> Statement verifying identity of above copies</p> <p><u>ACCOMPANYING APPLICATION PARTS</u></p> <p>7. <input type="checkbox"/> Assignment Papers (cover sheet & document(s))</p> <p>8. <input type="checkbox"/> 37 C.F.R. §3.73(b) Statement <input type="checkbox"/> Power of Attorney (when there is an assignee)</p> <p>9. <input type="checkbox"/> English Translation Document (if applicable)</p> <p>10. <input type="checkbox"/> Information Disclosure <input type="checkbox"/> Copies of IDS Citations Statement (IDS)PTO-1449</p> <p>11. <input type="checkbox"/> Preliminary Amendment</p> <p>12. <input checked="" type="checkbox"/> Return Receipt Postcard (MPEP 503) (should be specifically itemized)</p> <p>13. <input type="checkbox"/> *Small Entity <input type="checkbox"/> Statement filed in prior application Statement(s) (PTO/SB/09-12) Status still proper and desired</p> <p>14. <input type="checkbox"/> Certified Copy of Priority Document(s) (if foreign priority is claimed)</p> <p>15. <input checked="" type="checkbox"/> Other: Check for 1596.00</p>
--	--

16. If a CONTINUING APPLICATION, check appropriate box, and supply the requisite information below and in a preliminary amendment.

☐ Continuation ☐ Divisional ☐ Continuation-in-part (CIP) of prior application No.: _____

Prior application information: Examiner _____ Group/Art Unit: _____

For CONTINUATION or DIVISIONAL APPS. Only: The entire disclosure of the prior application, from which an oath or declaration is supplied under Box 4b, is considered to be part of the disclosure of the accompanying continuation or divisional application and is hereby incorporated by reference. The incorporation can only be relied upon when a portion has been inadvertently omitted from the submitted application parts.

17. CORRESPONDENCE ADDRESS

☐ Customer Number or Bar Code Label : : or ☒ Correspondence address below
:(Insert Customer No. Or Attach bar code label here) :

Name	COBRIN & GITTES				
Address	750 Lexington Avenue, 21 floor				
City	New York	State	New York	Zip Code	10022
Country	U S A.	Telephone	(212) 486-4000	Fax	(212) 486-4007

Name (Print/Type)	Lawrence E. Russ	Registration No. (Attorney/Agent)	35,342
Signature	<i>Lawrence E. Russ</i>	Date	January 28, 2000

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Box Patent Application, Washington, D.C. 20231.

EXPRESS MAIL CERTIFICATE

Date: 1/28/00

Label No. EL284833165US

I hereby certify that, on the date indicated above, I deposited this paper or fee with the U.S. Postal Service and that it was addressed for delivery to Box Patent Application, the Assistant Commissioner for Patents, Washington, D.C. 20231 by Express Mail Post Office to Addressee service.

Kevin Kinlaw
Name (Print)


Signature

METHOD OF AND APPARATUS FOR REDUCING ACOUSTIC NOISE IN WIRELESS AND LANDLINE BASED TELEPHONY

BACKGROUND OF THE INVENTION

The present invention is directed to wireless and landline based telephone communications and, more particularly, to reducing acoustic noise, such as background noise and system induced noise, present in wireless and landline based communication.

The perceived quality and intelligibility of speech transmitted over a wireless or landline based telephone lines is often degraded by the presence of background noise, coding noise, transmission and switching noise, etc. or by the presence of other interfering speakers and sounds. As an example, the quality of speech transmitted during a cellular telephone call may be affected by noises such as car engines, wind and traffic as well as by the condition of the transmission channel used.

Wireless telephone communication is also prone to providing lower perceived sound quality than wire based telephone communication because the speech coding process used during wireless communication removes a portion of the sound. Further, when the signal itself is noisy, the noise is encoded with the signal and further degrades the perceived sound quality because the speech coders used by these systems depend on encoding models intended for clean signals rather than for noisy signals. Wireless service providers, however, such as personal communication service (PCS) providers, attempt to deliver the same service and sound quality as landline telephony providers to attain greater consumer acceptance, and

therefore the PCS providers require improved end-to-end voice quality.

Additionally, transmitted noise degrades the capability of speech recognition systems used by various telephone services. The speech recognition systems are typically trained to recognize words or sounds under high transmission quality conditions and may fail to recognize words when noise is present.

In older wireline networks, such as are found in developing countries, system induced noise is often present because of poor wire shielding or the presence of cross talk which degrades sound quality. System induced noise is also present in more modern telephone communication systems because of the presence of channel static or quantization noise.

It is therefore desirable to provide wireless and landline telephone communication in which both the background noise and the system induced noise are reduced.

When noise reduction is carried out prior to encoding the transmitted signal, a significant portion of the additive noise is removed which results in better end-to-end perceived voice quality and robust speech coding. However, noise reduction is not always possible prior to encoding and therefore must be carried out after the signals have been received and/or decoded, such as at a base station or a switching center.

Existing commercial systems typically reduce encoded noise using spectral decomposition and spectral scaling. Known methods include estimating the noise level, computing the filter coefficients, smoothing the signal to noise ratio (SNR), and/or splitting the signal into respective bands. These methods, however, have the shortcomings that artifacts, known as musical noise, as well as speech distortions are produced.

Typically, the known noise reduction methods are based on generating an optimized filter that includes such methods as Wiener filtering, spectral subtraction and maximum

likelihood estimation. However, these methods are based on assumed idealized conditions that are rarely present during actual transmission. Additionally, these methods are not optimized for transmitting human speech or for human perception of speech, and therefore the methods must be altered for transmitting speech signals. Further, the conventional methods assume that the speech and noise spectra or the sub-band signal to noise ratio (SNR) are known beforehand, whereas the actual speech and noise spectra change over time and with transmission conditions. As a result, the band SNR is often incorrectly estimated and results in presence of musical noise. Additionally, when Wiener filtering is used, the filtering is based on minimum means square error (MMSE) optimized conditions that are not always appropriate for transmitting speech signals or for human perception of the speech signals.

Figure 1 illustrates a known method of spectral subtraction and scaling to filter noisy speech. A noisy speech signal is first buffered and windowed, as shown at step 102, and then undergoes a fast Fourier transform (FFT) into L frequency bins or bands, as shown at step 104. The energy of each of the bands is computed, as step 106 shows, and the noise level of each of the bands is estimated, as shown at step 110. The SNR is then estimated based on the computed energy and the estimated noise, as shown at step 108, and then a value of the filter gain is determined based on the estimated SNR, as shown at step 112. The calculated value of the gain is used as a multiplier value, as shown in step 114, and then the adjusted L frequency bins or bands undergo an inverse FFT or are passed through a synthesis filter bank, as step 116 shows, to generate an enhanced speech signal y_{bt} .

Various methods of carrying out the respective steps shown in figure 1 are known in the art:

As an example, U.S. Patent No. 4,811,404, titled "Noise Suppression System" to R.

Vimur et al. which issued on March 7, 1989, describes spectral scaling with sub-banding.

The spectral scaling is applied in a frequency domain using a FFT and an IFFT comprised of 128 speech samples or data points. The FFT bins are mapped into 16 non-homogeneous bands roughly following a known Bark scale.

5 When the filtered gains are computed for each sub-band, the amount of attenuation for each band is based on a non-linear function of the estimated SNR for that band. Bands having a SNR value less than 0 dB are assigned the lowest attenuation value of 0.17. Transient noise is detected based on the number of bands that are below or above the threshold value of 0 dB.

10 Noise energy values are estimated and updated during silent intervals, also known as stationary frames. The silent intervals are determined by first quantizing the SNR values according to a roughly exponential mapping and by then comparing the sum of the SNR values in 16 of the bands, known as a voice metric, to a threshold value. Alternatively, the noise energy value is updated using first-recursive averaging of the channel energy wherein an integration constant is based on whether the energy of a frame is higher than or similar to
15 the most recently estimated energy value.

20 Artifacts are removed by detecting very weak frames and then scaling these frames according the minimum gain value, 0.17. Sudden noise bursts in respective frames are detected by counting the number of bands in the frame whose SNR exceeds a predetermined threshold value. It is assumed that speech frames have a large number of bands that have a high SNR and that sudden noise burst is characterized by frames in which only a small number of bands have a high SNR.

Another example, European Patent No. EP 0,588,526 A1, titled "A Method Of And A

System For Noise Suppression" to Nokia Mobile Phones Ltd. which issued on March 23, 1994, describes using FFT for spectral analysis. Format locations are estimated whereby speech within the format locations is attenuated less than at other locations.

Noise is estimated only during speech intervals. Each of the filter passbands is split into two sub-bands using a special filter. The filter passbands are arranged such that one of the two sub-bands includes a speech harmonic and the other includes noise or other information and is located between two consecutive harmonic peaks.

Additionally, random flutter effect is avoided by not updating the filter coefficient during speech intervals. As a result, the filter gains convert poorly during changing noise and speech conditions.

A further example, U.S. Patent No. 5,485,522, titled "System For Adaptively Reducing Noise In Speech Signals" to T. Solve et al. which issued on January 16, 1996, is directed to attenuation applied in the time domain on the entire frame without sub-banding. The attenuation function is a logarithmic function of the noise level, rather than of the SNR, relative to a predefined threshold. When the noise level is less than the threshold, no attenuation is necessary. The attenuation function, however, is different when speech is detected in a frame rather than when the frame is purely noise.

A still further example, U.S. Patent No. 5,432,859, titled "Noise Reduction System" to J. Yang et al. which issued on July 11, 1995, describes using a sliding dual Fourier transform (DFT). Analysis is carried out on samples, rather than on frames, to avoid random fluctuation of flutter noise. An iterative expression is used to determine the DFT, and no inverse DFT is required. The filter gains of the higher frequency bins, namely those greater than 1 KHz, are set equal to the highest determined gain. The filter gains for the lower frequency bins are

calculated based on a known MMSE-based function of the SNR. When the SNR is less than -
6 dB, the gains are set to a predetermined small value.

It is desirable to provide noise reduction that avoids the weaknesses of the known
spectral subtraction and spectral scaling methods.

5

SUMMARY OF THE INVENTION

The present invention provides acoustic noise reduction for wireless or landline
telephony using frequency domain optimal filtering in which each frequency band of every
time frame is filtered as a function of the estimated signal-to-noise ratio (SNR) and the
estimated total noise energy for the frame and wherein non-speech bands, non-speech frames
and other special frames are further attenuated by one or more predetermined multiplier
values.

In accordance with the invention, noise in a transmitted signal comprised of frames
each comprised of frequency bands is reduced. A respective total signal energy and a
respective current estimate of the noise energy for at least one of the frequency bands is
determined. A respective local signal-to-noise ratio for at least one of the frequency bands is
determined as a function of the respective signal energy and the respective current estimate of
the noise energy. A respective smoothed signal-to-noise ratio is determined from the
respective local signal-to-noise ratio and another respective signal-to-noise ratio estimated for
a previous frame. A respective filter gain value is calculated for the frequency band from the
respective smoothed signal-to-noise ratio.

According to another aspect of the invention, noise is reduced in a transmitted signal.
It is determined whether at least a respective one as a plurality of frames is a non-speech

frame. When the frame is a non-speech frame, a noise energy level of at least one of the frequency bands of the frame is estimated. The band is filtered as a function of the estimated noise energy level.

Other features and advantages of the present invention will become apparent from the following detailed description of the invention with reference to the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention will now be described in greater detail in the following detailed description with reference to the drawings in which:

Figure 1 is a block diagram showing a known spectral subtraction scaling method.

Figure 2 is a block diagram showing a noise reduction method according to the invention.

Figure 3 shows the frames used to calculate the logarithm of the energy difference for detecting stationary frames.

Figures 4A and 4B show the filter coefficient values as a function of SNR for the known power subtraction filter and the Wiener filter and according to the invention.

Figure 5 shows the relation of the speech energy at the output of a noise reduction linear system according to the invention.

DETAILED DESCRIPTION OF THE INVENTION

The invention is an improvement of the known spectral subtraction and scaling method shown in Figure 1 and achieves better noise reduction with reduced artifacts by better estimating the noise level and by improved detection of non-speech frames. Additionally, the

invention includes a non-linear suppression scheme. Included are: (1) a new non-linear gain function that depends on the value of the smoothed SNR and which corrects the shortcomings of the Wiener filter and other classical filters that have a fast rising slope in the lower SNR region; (2) an adjustable aggressiveness control parameter that varies the percentage of the estimated noise that is to be removed (A set of spectral gains are derived based on the aggressiveness parameter and based on the nominal gain. The spectral gains are used to scale the FFT speech samples or points, and the nominal gains determine the feedback loop operation.); (3) non-speech frames are determined using at least one of four metrics: (a) a speech likelihood measure, (b) changes of the energy envelope, (c) a linear predictive coding (LPC) prediction error and (d) third order statistics of the LPC residual (Frames are determined to be non-speech frames when the signal is stationary for a predetermined interval. Stationary signals are detected as a function of changes in the energy envelope within a time window and based on the LPC prediction error. The LPC prediction error is used to avoid erroneously determining that frames representing sustained vowels or tones are non-speech frames. Alternatively, frames are determined to be non-speech frames based on the value of the normalized skewness of the LPC residual, namely the third order statistics of the LPC residual, and based on the LPC prediction error. As a further alternative, frames are determined to be non-speech frames based on the value of the frequency weighted noise likelihood measure determined across all frequency bands and combined with the LPC error.); (4) a "soft noise" estimation is used to determine the probability that a respective frame is noisy and is based on the log-likelihood measure; (5) a watchdog timer mechanism detects non-convergence of the updating of the estimated noise energy and forces an update when it times out (The forced update uses frames having a LPC prediction error outside the

nominal range for speech signals. The timer mechanism ensures proper convergence of the updated noise energy estimate and ensures fast updates.); and (6) marginal non-speech frames that are likely to contain only residual and musical noise are identified and further attenuated based on the total number of bands within the frame that have a high or low likelihood of representing speech signals, as well as based on the prediction error and the normalized skewness of the bands.

The invention carries out noise reduction processing in the frequency domain using a FFT and a perceptual band scale. In one example of the invention, the FFT speech samples or points are assigned to frequency bands along a perceptual frequency scale. Alternatively, frequency masking of neighboring speech samples carried out using a model of the auditory filters. Both methods attain noise reduction by filtering or scaling each frequency band based on a non-linear function of the SNR and other conditions.

Figure 2 is a block diagram showing the steps of a noise reduction method in accordance with the invention. The method is carried out iteratively over time. At each iteration, N new speech samples or points of noisy speech are read and combined with M speech samples from the preceding frame so that there is typically a 25% overlap between the new speech samples and those of the proceeding frame, though the actual percentage may be higher or lower. The combined frame is windowed and zero padded, as shown at step 202, and then a L point FFT is performed, as shown at step 204. Then, as shown at step 208, the squares of the real and imaginary components of the FFT are summed for each frequency point to attain the value of the signal energy $E_x(f)$. A local SNR, known as the SNR_{post} , is then calculated at each frequency point as the ratio of the total energy to the current estimate of the noise energy, as shown at step 208. The locally computed SNR is averaged with the

SNR estimated during the immediately preceding iteration of the filtering method, known as SNR_{est} , to obtain a smoothed SNR, as shown at step 214. The smoothed SNR is then used to compute the filter gains, as shown at step 210, which are applied to the FFT bins, as shown at step 216, and to compute the noise likelihood metric which are used to determine the speech and noise states, as step 232 shows. The filter gains are then used to calculate the value of the SNR_{est} for the next iteration.

To determine the value of the local SNR, the total energy and the current estimate of the noise energy are first convolved with the auditory filter centered at the respective frequency to account for frequency masking, namely the effective neighboring frequencies. The convolution operation results in a perceptual total energy value that is derived from the total signal energy $E_x(f)$ as follows:

$$E_x^p(f) = W(f) \otimes E_x(f) ,$$

where \otimes denotes convolution and $W(f)$ is the auditory filter centered at f . The convolution operation also results in a perceptual noise energy derived from the current estimate of the noise energy $E_n(f)$ as follows:

$$E_n^p(f) = W(f) \otimes E_n(f) .$$

Using the discrete value for the frequency, these relations become:

$$E_x^p(f) = \sum_{m=0}^{K-1} W\left(\frac{|f-m|}{f+0.5}\right) E_x(f), \text{ and}$$

$$E_n^p(f) = \sum_{m=0}^{K-1} W\left(\frac{|f-m|}{f+0.5}\right) E_n(f).$$

The local SNR at the frequency f is then determined from the relation:

$$SNR_{post}(f) = POS \left[\frac{E_x^p(f)}{E_n^p(f)} - 1 \right],$$

where the function $POS[x]$ has the value x when x is positive and has the value 0 otherwise.

The value SNR_{est} is then calculated from the relation:

$$SNR_{est}(f) = |G(f)|^2 \cdot SNR_{post}(f),$$

where the filter gains $G(s)$ are determined from the relation:

$$G(f) = C \cdot \sqrt{[SNR_{prior}(f)]}.$$

The values SNR_{post} and SNR_{est} are then averaged for the next iteration as follows:

$$SNR_{prior}(f) = (1 - \gamma)SNR_{post}(f) + \gamma SNR_{est}(f),$$

where the symbol γ is a smoothing constant having a value between 0.5 and 1.0 such that higher values of γ result in a smoother SNR.

15 The invention also detects the presence of non-speech frames by testing for of a stationary signal. The detection is based on changes in the energy envelope during a time interval and is based on the LPC prediction error. The log frame energy (FE), namely the logarithm of the sum of the signal energies for all frequency bands, is calculated for the current frame and for the previous K frames using the following relations:

$$FE_{dB} = 10 \cdot \log \left(\sum_f E_f \right).$$

The difference of the log frame energy is equivalent to determining the ratio of the energy between the current frame and each of the last K frames. The largest difference between the log frame energy of the current frame and that of each of the last K frames is determined, as shown in Figure 3. When the largest difference is less than a predefined threshold value, the energy contour has not changed over the interval of the K frames, and thus the signal is stationary.

When the largest difference exceeds the threshold value for a preset time period, known as a hangover period, the stationary frames are likely to be non-speech frames because speech utterances typically have changing energy contours within time intervals of 0.5 to 1 seconds. However, the signal may be stationary signal during the utterance of a sustained vowel or during the presence of a in-band tone, such as a dial tone. To eliminate the likelihood of falsely detecting a non-speech frame, an LPC prediction error, which is the inverse of the LPC prediction gain, is determined from the reflection coefficient generated by the LPC analysis performed at the speech encoder. The LPC prediction error (PE) is determined from the following relation:

$$PE = \prod_{k=0}^{K-1} [1 - r_{ck}^2].$$

A low prediction error indicates the presence of speech frames, a near zero prediction error indicates the presence of sustained vowels or in-band tones, and a high prediction error

indicates the presence of non-speech frames.

When the LPC prediction error is greater than a preset threshold value and the change of the log frame energies over the preceding K frames is less than another threshold value, a stationarity counter is activated and remains active up to the duration of the hangover period.

- 5 When the stationarity counter reaches a preset value, the frame is determined to be stationary.

Figure 2 also shows the detection of stationary frames by computing the LPC error, as shown at step 220, and the determination of stationarity, as step 222 shows. The log frame energies of the proceeding K frames is determined from the energy values determined at step 206.

The invention also determines the presence of non-speech frames using a statistical speech likelihood measurement from all the frequency bands of a respective frame. For each of the bands, the likelihood measure, $\Lambda(f)$, is determined from the local SNR and the smoothed SNR described above using the following relation:

15
$$\Lambda(f) = \frac{e^{\left[\left(\frac{SNR_{prior}(f)}{1 + SNR_{prior}(f)} \right) SNR_{post}(f) \right]}}{1 + SNR_{prior}(f)}.$$

The above relation is derived from a known statistical model for determining the FFT magnitude for speech and noise signals.

In accordance with the invention, the statistical speech likelihood measure of each

frequency band is weighted by a frequency weighting function prior to combining the log frame likelihood measure across all the frequency bands. The weighting function accounts for the distribution of speech energy across the frequencies and for the sensitivity of human hearing as a function of the frequency. The weighted values are combined across all bands to produce a frame likelihood metric shown by the following relation:

$$NoiseLikelihood = \sum_f W(f) \cdot \log[\Lambda(f)].$$

To prevent the false detection of low amplitude speech segments, the noise likelihood is combined with the LPC prediction error described above before a decision is made to determine whether the frame is non-speech.

The invention also determines whether a frame is non-speech based on the normalized skewness of the LPC residual, namely based on the third order statistics of the sampled LPC residual $e(n)$, $E[e(n)^3]$, which has a non-zero value for speech signals and has a value of zero in the presence of Gaussian noise. The skewness is typically normalized either by its variance, which is a function of the frame length, or by the estimate of the noise energy. The energy of the LPC residual, E_x , is determined from the following relation:

$$E_x = \frac{1}{N} \sum_{n=0}^{N-1} [e(n)]^2.$$

where $e(n)$ are the sampled values of the LPC residual, and N is the frame length. The skewness SK of the LPC residual is determined as follows:

$$SK = \frac{1}{N} \sum_{n=0}^{N-1} [e(n)]^3.$$

The value of the normalized skewness as a function of the total energy is then determined from the following relation:

$$\gamma_3 = \frac{SK}{E_x^{1.5}}.$$

For a Gaussian process, the variance of the skewness has the following relation:

$$Var[SK] = \frac{15E_n^3}{N},$$

where E_n is the estimate of the noise energy. The normalized skewness based on the variance of the skewness is determined from the following relation:

$$\gamma_3' = \frac{SK}{\sqrt{\frac{15E_n^3}{N}}}.$$

To detect the presence of non-speech frames, both the normalized skewness and the skewness combined with the LPC prediction error are utilized, as shown in Table 1.

Whenever a frame is determined to be a non-speech frame based on any of the above three methods, an updated noise energy value is estimated. Also, when the current estimate

of the noise energy of a band in a frame is greater than the total energy of the band, the updated noise energy is similarly estimated. The estimated noise energy is updated by a smoothing operation in which the value of a smoothing constant depends on the condition required for estimating the noise energy. The new estimated noise energy value $E(m+1, f)$ of each frequency band of a frame is determined from the prior estimated value $E(m, f)$ and from the band energy $E_{ch}(m, f)$ using the following relation:

$$E(m+1, f) = (1 - \alpha)E(m, f) + \alpha E_{ch}(m, f)$$

where m is the iteration index and α is the update constant.

The estimation of the noise energy is essentially a feedback loop because the noise energy is estimated during non-speech intervals and is detected based on values such as the SNR and the normalized skewness which are, in turn, functions of previously estimated noise energy values. The feedback loop may fail to converge when, for example, the noise energy level goes to near zero for an interval and then again increases. This situation may occur, for example, during a cellular telephone handoff where the signal received from the mobile phone drops to zero at the base station for a short time period, typically about a second, and then again rises. Typically, the normalized skewness value, which is based on third order statistics, is not affected by such changes in the estimated noise level. However, the third order statistics do not always prevent failure to converge.

Therefore, the invention includes a watch dog timer to monitor the convergence of the noise estimation feed back loop by monitoring the time that has elapsed from the last noise energy update. If the estimated noise energy has not been updated within a preset time-out interval, typically three seconds, it is assumed that the feedback loop is not converging, and a forced noise energy value is used to return the feedback loop back to operation. Because a

forced estimated noise energy update is used, the corresponding speech frame is not used and, instead, the LPC prediction error is used to select the next frame or frames having a sufficiently high prediction error and therefore reduce the likelihood of any subsequent failures to converge. A forced update condition may continue as long as the feedback loop fails to converge. Typically, the duration of the forced update needed to bring the feedback loop back in convergence is fewer than five frames.

Table 1 shows the conditions for which the estimated noise energy is forcibly updated and shows the value of the update constant α corresponding to a respective condition. When the watch dog timer has expired, the update constant has a value of 0.002. When a frame is determined to be stationary, the update constant has a value of 0.05. When the noise likelihood is less than a threshold value T_{LIK} and the LPC prediction error is greater than a threshold value T_{PE2} , the update constant has a value of 0.1. When the normalized skewness of the LPC residual has a near-zero value, namely when it has an absolute value less than a threshold T_a (when normalized by total energy) or less than T_b (when normalized by the variance), and when the LPC prediction error is greater than a threshold value T_{PE2} , the update constant has a value of 0.05. When the current noise energy estimate is greater than the total energy, namely when the noise energy is decreasing, the update constant has a value of 0.1.

The invention also provides a filter gain function that reaches unity for SNR values above 13 dB, as Figures 4A and 4B show. At these values, the speech sounds mask the noise so that no attenuation is needed. Known classical filters, such as the Wiener filter or the power subtraction filter, have a filter gain function that rises quickly in the region where the SNR is just below 10dB. The rapid rise in filter gain causes fluctuations in the output amplitude of the speech signals.

The gain function of the invention provides for a more slowly rising filter gain in this region so that the filter gain reaches a value of unity for SNR values above 13dB. The smoothed SNR, SNR_{prior} , is used to determine the gain function, rather than the value of the local SNR, SNR_{post} , because the local SNR is found to behave more erratically during non-
5 speech and weak-speech frames. The filter gain function is therefore determined by the following relation:

$$G(f) = C \cdot \sqrt{[SNR_{prior}(f)]} ,$$

where C is a constant that controls the steepness of the rise of the gain function and has a value between 0.15 and 0.25 and depends on the noise energy.

Further, when the speech likelihood metric described above is less than the speech threshold value, namely when the frequency band is likely to be comprised only of noise, the gain function $G(f)$ is forced to have a minimum gain value. The gain values are then applied to the FFT frequency bands, as shown at step 216 of Figure 2, prior to carrying out the IFFT, as shown at step 240.

The invention also provides for further control of the filter gains using a control parameter F, known as the aggressiveness "knob", that further controls the amount of noise removed and which has a value between 0 and 1. The aggressiveness knob parameter allows for additional control of the noise reduction and prevents distortion that results from the excessive removal of noise. Modified filter gains $G'(f)$ are then determined from the above
15 filter gains $G(f)$ and from the aggressiveness knob parameter F according to the following relation:

$$G'(f) = \sqrt{[1 - F \cdot (1 - G(f)^2)]} .$$

The modified gain values are then applied to the corresponding FFT sample values in the manner described above.

The value of the aggressiveness knob parameter F may also vary with the frequency band of the frame. As an example, band having a frequencies less than 1kHz may have high aggressiveness, namely high F values, because these bands have high speech energy, whereas bands having frequencies between 1 and 3 kHz may have a lower value of F.

Figure 5 shows the relation between the input and output energies of the speech bands as a function of the filter gain. The speech energy at the output of the suppression filter 502 is determined from the following relation:

$$E_x = |G(f)|^2 \cdot E_x .$$

The noise energy removed is the difference between the output energy and the input energy and is shown as follows:

$$E_n = E_x - |G(f)|^2 \cdot E_x$$

However, with certain frequencies, the removal of only a fraction of the noise is desirable.

When the noise energy that is removed is adjusted based on the aggressiveness knob parameter F, the following relation is used:

$$E_n' = E_x - |G'(f)|^2 \cdot E_x = F \{E_x - |G(f)|^2 \cdot E_x\}$$

From this relation, the above equation determining the value of the adjusted gain G'(f) is derived.

The invention also detects and attenuates frames consisting solely of musical noise bands, namely frames in which a small percentage of the bands have a strong signal that, after

processing, generates leftover noise having sounds similar to musical sounds. Because such frames are non-speech frames, the normalized skewness of the frame will not exceed its threshold value and the LPC prediction error will not be less than its threshold value so that the musical noise cannot ordinarily be detected. To detect these frames, the number of frequency bands having a likelihood metric above a threshold value are counted, the threshold value indicating that the bands are strong speech bands, and when the strong speech bands are less than 25% of the total number of frequency bands, the strong speech bands are likely to be musical noise bands and not actual speech bands. The detected speech bands are further attenuated by setting the filter gains $G(f)$ of the frame to its minimum value.

Although the present invention has been described in relation to particular embodiment thereof, many other variations and modifications and other uses may become apparent to those skilled in the art. It is preferred, therefore, that the present invention be limited not by the specific disclosure herein, but only by the appended claims.

WHAT IS CLAIMED IS:

1. A method of reducing noise in a transmitted signal comprised of a plurality of frames, each of said frames including a plurality of frequency bands; said method comprising the steps of :

determining a respective total signal energy and a respective current estimate of the noise energy for at least one of said plurality of frequency bands of at least one of said plurality of frames;

determining a respective local signal-to-noise ratio (SNR_{post}) for said at least one of said plurality of frequency bands as a function of said respective signal energy and said respective current estimate of the noise energy;

determining a respective smoothed signal-to-noise ratio (SNR_{prior}) for said at least one of said plurality of frequency bands from said respective local signal-to-noise ratio and another respective signal-to-noise ratio (SNR_{est}) estimated for a previous frame; and

calculating a respective filter gain value for said at least one of said plurality of frequency bands from said respective smoothed signal-to-noise ratio.

2. The method of claim 1 wherein said respective local signal-to-noise ratio (SNR_{post}) is determined by the following relation:

$$SNR_{post}(f) = POS \left[\frac{E_x^p(f)}{E_n^p(f)} - 1 \right],$$

wherein $POS[x]$ has the value x when x is positive and has the value 0 otherwise, $E_x(f)$ is said respective total signal energy and $E_n(f)$ is said respective current estimate of the noise energy.

3. The method of claim 1 wherein said estimated respective signal-to-noise ratio (SNR_{est}) is determined by the following relation:

1870
 1871
 1872
 1873
 1874
 1875
 1876
 1877
 1878
 1879
 1880
 1881
 1882
 1883
 1884
 1885
 1886
 1887
 1888
 1889
 1890
 1891
 1892
 1893
 1894
 1895
 1896
 1897
 1898
 1899
 1900
 1901
 1902
 1903
 1904
 1905
 1906
 1907
 1908
 1909
 1910
 1911
 1912
 1913
 1914
 1915
 1916
 1917
 1918
 1919
 1920
 1921
 1922
 1923
 1924
 1925
 1926
 1927
 1928
 1929
 1930
 1931
 1932
 1933
 1934
 1935
 1936
 1937
 1938
 1939
 1940
 1941
 1942
 1943
 1944
 1945
 1946
 1947
 1948
 1949
 1950
 1951
 1952
 1953
 1954
 1955
 1956
 1957
 1958
 1959
 1960
 1961
 1962
 1963
 1964
 1965
 1966
 1967
 1968
 1969
 1970
 1971
 1972
 1973
 1974
 1975
 1976
 1977
 1978
 1979
 1980
 1981
 1982
 1983
 1984
 1985
 1986
 1987
 1988
 1989
 1990
 1991
 1992
 1993
 1994
 1995
 1996
 1997
 1998
 1999
 2000
 2001
 2002
 2003
 2004
 2005
 2006
 2007
 2008
 2009
 2010
 2011
 2012
 2013
 2014
 2015
 2016
 2017
 2018
 2019
 2020
 2021
 2022
 2023
 2024
 2025
 2026
 2027
 2028
 2029
 2030
 2031
 2032
 2033
 2034
 2035
 2036
 2037
 2038
 2039
 2040
 2041
 2042
 2043
 2044
 2045
 2046
 2047
 2048
 2049
 2050
 2051
 2052
 2053
 2054
 2055
 2056
 2057
 2058
 2059
 2060
 2061
 2062
 2063
 2064
 2065
 2066
 2067
 2068
 2069
 2070
 2071
 2072
 2073
 2074
 2075
 2076
 2077
 2078
 2079
 2080
 2081
 2082
 2083
 2084
 2085
 2086
 2087
 2088
 2089
 2090
 2091
 2092
 2093
 2094
 2095
 2096
 2097
 2098
 2099
 2100
 2101
 2102
 2103
 2104
 2105
 2106
 2107
 2108
 2109
 2110
 2111
 2112
 2113
 2114
 2115
 2116
 2117
 2118
 2119
 2120
 2121
 2122
 2123
 2124
 2125
 2126
 2127
 2128
 2129
 2130
 2131
 2132
 2133
 2134
 2135
 2136
 2137
 2138
 2139
 2140
 2141
 2142
 2143
 2144
 2145
 2146
 2147
 2148
 2149
 2150
 2151
 2152
 2153
 2154
 2155
 2156
 2157
 2158
 2159
 2160
 2161
 2162
 2163
 2164
 2165
 2166
 2167
 2168
 2169
 2170
 2171
 2172
 2173
 2174
 2175
 2176
 2177
 2178
 2179
 2180
 2181
 2182
 2183
 2184
 2185
 2186
 2187
 2188
 2189
 2190
 2191
 2192
 2193
 2194
 2195
 2196
 2197
 2198
 2199
 2200
 2201
 2202
 2203
 2204
 2205
 2206
 2207
 2208
 2209
 2210
 2211
 2212
 2213
 2214
 2215
 2216
 2217
 2218
 2219
 2220
 2221
 2222
 2223
 2224
 2225
 2226
 2227
 2228
 2229
 2230
 2231
 2232
 2233
 2234
 2235
 2236
 2237
 2238
 2239
 2240
 2241
 2242
 2243
 2244
 2245
 2246
 2247
 2248
 2249
 2250
 2251
 2252
 2253
 2254
 2255
 2256
 2257
 2258
 2259
 2260
 2261
 2262
 2263
 2264
 2265
 2266
 2267
 2268
 2269
 2270
 2271
 2272
 2273
 2274
 2275
 2276
 2277
 2278
 2279
 2280
 2281
 2282
 2283
 2284
 2285
 2286
 2287
 2288
 2289
 2290
 2291
 2292
 2293
 2294
 2295
 2296
 2297
 2298
 2299
 2300
 2301
 2302
 2303
 2304
 2305
 2306
 2307
 2308
 2309
 2310
 2311
 2312
 2313
 2314
 2315
 2316
 2317
 2318
 2319
 2320
 2321
 2322
 2323
 2324

5 current estimate of the noise energy level of said at least one of said plurality of bands of said at least one of said plurality of frames; and

determining said respective filter gain value as a function of said updated current estimate of the noise energy level.

9. The method of claim 8 wherein said at least one of said plurality of frames is determined to be a non-speech frame when said at least one frame is a stationary frame.

10. The method of claim 9 wherein said at least a respective one of said plurality of frames is determined to be a stationary frame when a difference in a logarithm of an energy of said at least one frame and a logarithm in an energy of at a prior one of said plurality of frames exceeds a predefined threshold value.

11. The method of claim 8 wherein said at least one of said plurality of frames is determined to be a non-speech frame as a function of a sum of weighted values each corresponding to a respective one of said frequency bands of said respective one of said plurality of frames, each of said weighted values being a product of a logarithm of a speech likelihood metric of said corresponding one of said frequency bands and a weighting factor of said corresponding one of said frequency bands, and as a function of a linear predictive coding (LPC) prediction error.

12. The method of claim 11 wherein said speech likelihood metric of said corresponding one of said frequency bands is determined by the following relation:

$$\Lambda(f) = \frac{e^{\left[\frac{SNR_{prior}(f)}{1 + SNR_{prior}(f)} \right] SNR_{post}(f)}}}{1 + SNR_{prior}(f)}, \text{ wherein } SNR_{post} \text{ is said respective local signal-}$$

to-noise ratio and SNR_{prior} is said respective smoothed signal-to-noise ratio.

13. The method of claim 8 wherein said at least a respective one of said plurality of frames is determined to be a non-speech frame as a function of a normalized skewness value of a linear predictive coding (LPC) residual of said at least a respective one of said plurality of frames and as a function of a linear predictive coding (LPC) prediction error.

14. The method of claim 13 wherein said skewness value of said LPC residual is determined by the following relation:

$$SK = \frac{1}{N} \sum_{n=0}^{N-1} [e(n)]^3, \text{ wherein } e(n) \text{ are sampled values of an LPC residual, and } N \text{ is a frame}$$

length.

15. The method of claim 14 wherein said skewness value is normalized by a function of an estimated value of a total energy of said respective one of said plurality of frames, said total energy being determined by the following relation:

$$E_x = \frac{1}{N} \sum_{n=0}^{N-1} [e(n)]^2, \text{ wherein } e(n) \text{ are sampled values of an LPC residual, and } N \text{ is a frame}$$

length.

16. The method of claim 14 wherein said skewness value is normalized by an estimated value of a variance of said skewness value, said variance being determined by the following relation:

$$Var[SK] = \frac{15E_n^3}{N}, \text{ wherein } E_n \text{ is said current estimate of the noise energy level and } N \text{ is a}$$

frame length.

17. The method of claim 8 wherein said current estimate of the noise energy level is determined by the following relation:

$E(m+1, f) = (1 - \alpha)E(m, f) + \alpha E_{ch}(m, f)$, wherein $E(m, f)$ is a prior estimated noise energy level, $E_{ch}(m, f)$ is a band energy, m is an iteration index and α is an update constant.

18. The method of claim 17 wherein a value of said update constant α is determined by one of a watchdog timer being expired, said at least one of said plurality of frames being stationary, said at least one of said plurality of frames being a non-speech frame, a LPC residual of said at least one of said plurality of frames having substantially zero skewness, and a current value of said estimated noise energy level being greater than a total energy of said plurality of frames.

19. The method of claim 17 wherein said estimated noise level is forced to be updated using a noise energy level of a current frame when said estimated noise level is not updated within a preset interval.

20. The method of claim 1 wherein said filtering gain is further adjusted as a function of an aggressiveness setting parameter (F) according to the following relation:

$G'(f) = \sqrt{[1 - F \cdot (1 - G(f)^2)]}$, wherein $G(f)$ is said filtering gain prior to being adjusted.

21. The method of claim 1 further comprising the steps of: determining a respective speech likelihood metric of each of said plurality of said frequency bands of said at least one of said plurality of frames; determining a number of said plurality of said frequency bands having said respective speech likelihood metric above a threshold value; and setting, when said number exceeds a predetermined percentage of a total number of said plurality of

said frequency bands, said filter gain for each of said plurality of said frequency bands to a minimum value.

22. The method of claim 11 wherein an said filtering gain is set to a minimum value when said speech likelihood metric is less than a threshold value.

23. A method of reducing noise in a transmitted signal comprised of a plurality of frames, each of said frames including a plurality of frequency bands; said method comprising the steps of :

determining whether at least a respective one of said plurality of frames is a non-
5 speech frame;

estimating, when said at least one of said plurality of frames is a non-speech frame, a noise energy level of at least one of said plurality of bands of said at least a respective one of said plurality of frames; and

filtering said at least one band as a function of said estimated noise level.

24. The method of claim 23 wherein said at least a respective one of said plurality of frames is determined to be a non-speech frame when said at least one frame is a stationary frame.

25. The method of claim 24 wherein said at least a respective one of said plurality of frames is determined to be a stationary frame when a difference in a logarithm of an energy of said at least one frame and a logarithm in an energy of at a prior one of said plurality of frames exceeds a predefined threshold value.

26. The method of claim 23 wherein said at least a respective one of said plurality of frames is determined to be a non-speech frame as a function of a sum of weighted values each corresponding to a respective one of said frequency bands of said respective one of said

plurality of frames, each of said weighted values being a product of a logarithm of a speech
 5 likelihood metric of said corresponding one of said frequency bands and a weighting factor of
 said corresponding one of said frequency bands, and as a function of a linear predictive
 coding (LPC) prediction error.

27. The method of claim 26 wherein said speech likelihood metric of said
 corresponding one of said frequency bands is determined by the following relation:

$$\Lambda(f) = \frac{e^{\left[\frac{SNR_{prior}(f)}{1 + SNR_{prior}(f)} \right] SNR_{post}(f)}}}{1 + SNR_{prior}(f)}, \text{ wherein } SNR_{post} \text{ is said respective local signal-}$$

to-noise ratio and SNR_{prior} is said respective smoothed signal-to-noise ratio.

28. The method of claim 23 wherein said at least a respective one of said plurality
 of frames is determined to be a non-speech frame as a function of a normalized skewness
 value of a linear predictive coding (LPC) residual of said at least a respective one of said
 plurality of frames and as a function of a linear predictive coding (LPC) prediction error.

29. The method of claim 28 wherein said skewness value of said LPC residual is
 determined by the following relation:

$$SK = \frac{1}{N} \sum_{n=0}^{N-1} [e(n)]^3, \text{ wherein } e(n) \text{ are sampled values of an LPC residual, and } N \text{ is a frame}$$

length.

30. The method of claim 29 wherein said skewness value is normalized by a function of an estimated value of a total energy of said respective one of said plurality of frames, said total energy being determined by the following relation:

$$E_x = \frac{1}{N} \sum_{n=0}^{N-1} [e(n)]^2, \text{ wherein } e(n) \text{ are sampled values of an LPC residual, and } N \text{ is a frame}$$

length.

31. The method of claim 29 wherein said skewness value is normalized by an estimated value of a variance of said skewness value, said variance being determined by the following relation:

$$Var[SK] = \frac{15E_n^3}{N}, \text{ wherein } E_n \text{ is said current estimate of the noise energy level and } N \text{ is a}$$

frame length.

32. The method of claim 23 wherein said estimated noise level is determined by the following relation:

$$E(m+1, f) = (1 - \alpha)E(m, f) + \alpha E_{ch}(m, f), \text{ wherein } E(m, f) \text{ is a prior estimated noise energy level, } E_{ch}(m, f) \text{ is a band energy, } m \text{ is an iteration index and } \alpha \text{ is an update constant.}$$

33. The method of claim 32 wherein a value of said update constant α is determined by one of a watchdog timer being expired, said at least one of said plurality of frames being stationary, said at least one of said plurality of frames being a non-speech frame, a LPC residual of said at least one of said plurality of frames having substantially zero skewness, and a current value of said estimated noise energy level being greater than a total

energy of said plurality of frames.

34. An apparatus of reducing noise in a transmitted signal including a plurality of frames, each of said frames including a plurality of frequency bands; said apparatus comprising:

means for determining a respective total signal energy and a respective current estimate of the noise energy for at least one of said plurality of frequency bands of at least one of said plurality of frames;

means for determining a respective local signal-to-noise ratio (SNR_{post}) for said at least one of said plurality of frequency bands as a function of said respective signal energy and said respective current estimate of the noise energy;

means for determining a respective smoothed signal-to-noise ratio (SNR_{prior}) for said at least one of said plurality of frequency bands from said respective local signal-to-noise ratio and another respective signal-to-noise ratio (SNR_{est}) estimated for a previous frame; and

means for calculating a respective filter gain value for said at least one of said plurality of frequency bands from said respective smoothed signal-to-noise ratio.

35. The apparatus of claim 34 wherein said respective local signal-to-noise ratio (SNR_{post}) is determined by the following relation:

$$SNR_{post}(f) = POS \left[\frac{E_s^p(f)}{E_n^p(f)} - 1 \right],$$

wherein $POS[x]$ has the value x when x is positive and has the value 0 otherwise, $E_s(f)$ is said respective total signal energy and $E_n(f)$ is said respective current estimate of the noise energy.

36. The apparatus of claim 34 wherein said estimated respective signal-to-noise ratio (SNR_{est}) is determined by the following relation:

$$SNR_{est}(f) = |G(f)|^2 \cdot SNR_{post}(f) ,$$

wherein $G(f)$ is a prior respective signal gain and SNR_{post} is said respective local signal-to-noise ratio.

37. The apparatus of claim 34 wherein said respective smoothed signal-to-noise ratio (SNR_{prior}) is determined by the following relation:

$$SNR_{prior}(f) = (1 - \gamma) SNR_{post}(f) + \gamma SNR_{est}(f),$$

wherein γ is a smoothing constant, SNR_{post} is said respective local signal-to-noise ratio and SNR_{est} is said estimated respective signal-to-noise ratio.

38. The apparatus of claim 34 wherein said respective filter gain value is determined by the following relation:

$$G(f) = C \cdot \sqrt{[SNR_{prior}(f)]} ,$$

wherein SNR_{prior} is said respective smoothed signal-to-noise ratio.

39. The apparatus of claim 34 further comprising the means for forming said at least one of said plurality of frames from a first number of new speech samples and a second number of prior speech samples.

40. The apparatus of claim 34 further comprising means for forming said plurality of frequency bands by carrying out a fast Fourier transform (FFT) on said at least one of said plurality of frames.

41. The apparatus of claim 34 further comprising:

means for determining whether said at least one of said plurality of frames is a non-speech frame;

means for updating, when said at least one of said plurality of frames is a non-speech

5 frame, said current estimate of the noise energy level of said at least one of said plurality of bands of said at least one of said plurality of frames; and

means for determining said respective filter gain value as a function of said updated current estimate of the noise energy level.

42. The apparatus of claim 41 wherein said at least one of said plurality of frames is determined to be a non-speech frame when said at least one frame is a stationary frame.

43. The apparatus of claim 42 wherein said at least a respective one of said plurality of frames is determined to be a stationary frame when a difference in a logarithm of an energy of said at least one frame and a logarithm in an energy of at a prior one of said plurality of frames exceeds a predefined threshold value.

44. The apparatus of claim 42 wherein said at least one of said plurality of frames is determined to be a non-speech frame as a function of a sum of weighted values each corresponding to a respective one of said frequency bands of said respective one of said plurality of frames, each of said weighted values being a product of a logarithm of a speech
5 likelihood metric of said corresponding one of said frequency bands and a weighting factor of said corresponding one of said frequency bands, and as a function of a linear predictive coding (LPC) prediction error.

45. The apparatus of claim 44 wherein said speech likelihood metric of said corresponding one of said frequency bands is determined by the following relation:

$$\Lambda(f) = \frac{e^{\left[\left(\frac{SNR_{prior}(f)}{1 + SNR_{prior}(f)} \right) SNR_{post}(f) \right]}}{1 + SNR_{prior}(f)}, \text{ wherein } SNR_{post} \text{ is said respective local signal-}$$

to-noise ratio and SNR_{prior} is said respective smoothed signal-to-noise ratio.

46. The apparatus of claim 41 wherein said at least a respective one of said plurality of frames is determined to be a non-speech frame as a function of a normalized skewness value of a linear predictive coding (LPC) residual of said at least a respective one of said plurality of frames and as a function of a linear predictive coding (LPC) prediction error.

47. The apparatus of claim 46 wherein said skewness value of said LPC residual is determined by the following relation:

$$SK = \frac{1}{N} \sum_{n=0}^{N-1} [e(n)]^3, \text{ wherein } e(n) \text{ are sampled values of an LPC residual, and } N \text{ is a frame}$$

length.

48. The apparatus of claim 47 wherein said skewness value is normalized by an estimated value of a total energy of said respective one of said plurality of frames, said total energy being determined by the following relation:

$$E_x = \frac{1}{N} \sum_{n=0}^{N-1} [e(n)]^2, \text{ wherein } e(n) \text{ are sampled values of an LPC residual, and } N \text{ is a frame}$$

length.

49. The apparatus of claim 47 wherein said skewness value is normalized by an estimated value of a variance of said skewness value, said variance being determined by the following relation:

$$Var[SK] = \frac{15E_n^3}{N}, \text{ wherein } E_n \text{ is said current estimate of the noise energy level and } N \text{ is a}$$

frame length.

50. The apparatus of claim 41 wherein said estimated noise level is determined by the following relation:

$E(m+1, f) = (1 - \alpha)E(m, f) + \alpha E_{ch}(m, f)$, wherein $E(m, f)$ is a prior estimated noise energy level, $E_{ch}(m, f)$ is a band energy, m is an iteration index and α is an update constant.

51. The apparatus of claim 50 wherein a value of said update constant α is determined by one of a watchdog timer being expired, said at least one of said plurality of frames being stationary, said at least one of said plurality of frames being a non-speech frame, a LPC residual of said at least one of said plurality of frames having substantially zero skewness, and a current value of said estimated noise energy level being greater than a total energy of said plurality of frames.

52. The apparatus of claim 41 wherein said estimated noise level is forced to be updated using a noise energy level of a current frame when said estimated noise level is not updated within a preset interval.

53. The apparatus of claim 34 wherein said filtering gain is further adjusted as a function of an aggressiveness setting parameter (F) according to the following relation:

$G'(f) = \sqrt{[1 - F \cdot (1 - G(f)^2)]}$, wherein $G(f)$ is said filtering gain prior to being adjusted.

54. The apparatus of claim 34 further comprising the steps of: determining a respective speech likelihood metric of each of said plurality of said frequency bands of said at least one of said plurality of frames; determining a number of said plurality of said frequency bands having said respective speech likelihood metric above a threshold value; and setting, when said number exceeds a predetermined percentage of a total number of said plurality of

said frequency bands, said filter gain for each of said plurality of said frequency bands to a minimum value.

55. The apparatus of claim 44 wherein said filtering gain is set to a minimum value when said speech likelihood metric is less than a threshold value.

56. An apparatus of reducing noise in a transmitted signal including a plurality of frames, each of said frames including a plurality of frequency bands; said apparatus comprising the steps of :

means for determining whether at least a respective one of said plurality of frames is a
5 non-speech frame;

means for estimating, when said at least one of said plurality of frames is a non-speech frame, a noise energy level of at least one of said plurality of bands of said at least a respective one of said plurality of frames; and

means for filtering said at least one band as a function of said estimated noise level.

57. The apparatus of claim 56 wherein said at least a respective one of said plurality of frames is determined to be a non-speech frame when said at least one frame is a stationary frame.

58. The apparatus of claim 57 wherein said at least a respective one of said plurality of frames is determined to be a stationary frame when a difference in a logarithm of an energy of said at least one frame and a logarithm in an energy of at a prior one of said plurality of frames exceeds a predefined threshold value.

59. The apparatus of claim 56 wherein said at least a respective one of said plurality of frames is determined to be a non-speech frame as a function of a sum of weighted values each corresponding to a respective one of said frequency bands of said respective one

of said plurality of frames, each of said weighted values being a product of a logarithm of a
 5 speech likelihood metric of said corresponding one of said frequency bands and a weighting
 factor of said corresponding one of said frequency bands, and as a function of a linear
 predictive coding (LPC) prediction error.

60. The apparatus of claim 59 wherein said speech likelihood metric of said
 corresponding one of said frequency bands is determined by the following relation:

$$\Lambda(f) = \frac{e^{\left[\frac{SNR_{prior}(f)}{1 + SNR_{prior}(f)} \right] SNR_{post}(f)}}}{1 + SNR_{prior}(f)}, \text{ wherein } SNR_{post} \text{ is said respective local signal-}$$

to-noise ratio and SNR_{prior} is said respective smoothed signal-to-noise ratio.

61. The apparatus of claim 56 wherein said at least a respective one of said
 plurality of frames is determined to be a non-speech frame as a function of a normalized
 skewness value of a linear predictive coding (LPC) residual of said at least a respective one of
 said plurality of frames and as a function of a linear predictive coding (LPC) prediction error.

62. The apparatus of claim 61 wherein said skewness value of said LPC residual is
 determined by the following relation:

$$SK = \frac{1}{N} \sum_{n=0}^{N-1} [e(n)]^3, \text{ wherein } e(n) \text{ are sampled values of an LPC residual, and } N \text{ is a frame}$$

length.

63. The apparatus of claim 61 wherein said skewness value is normalized by an estimated value of a total energy of said respective one of said plurality of frames, said total energy being determined by the following relation:

$$E_x = \frac{1}{N} \sum_{n=0}^{N-1} [e(n)]^2, \text{ wherein } e(n) \text{ are sampled values of an LPC residual, and } N \text{ is a frame}$$

length.

64. The apparatus of claim 62 wherein said skewness value is normalized by an estimated value of a variance of said skewness value, said variance being determined by the following relation:

$$Var[SK] = \frac{15E_n^3}{N}, \text{ wherein } E_n \text{ is said current estimate of the noise energy level and } N \text{ is a}$$

frame length.

65. The apparatus of claim 56 wherein said estimated noise level is determined by the following relation:

$$E(m+1, f) = (1 - \alpha)E(m, f) + \alpha E_{ch}(m, f), \text{ wherein } E(m, f) \text{ is a prior estimated noise energy level, } E_{ch}(m, f) \text{ is a band energy, } m \text{ is an iteration index and } \alpha \text{ is an update constant.}$$

66. The apparatus of claim 65 wherein a value of said update constant α is determined by one of a watchdog timer being expired, said at least one of said plurality of frames being stationary, said at least one of said plurality of frames being a non-speech frame, a LPC residual of said at least one of said plurality of frames having substantially zero skewness, and a current value of said estimated noise energy level being greater than a total

energy of said plurality of frames.

ABSTRACT OF THE DISCLOSURE

Acoustic noise for wireless or landline telephony is reduced using frequency domain of optimal filtering in which each frequency band of every time frame is filtered as a function of the estimated signal-to-noise ratio and the estimated total noise energy for the frame. Non-speech, non-speech frames and other special frames are further attenuated by one or more predetermined multiplier values. Noise in a transmitted signal comprised of frames each comprised of frequency bands is reduced. A respective total signal energy and a respective current estimate of the noise energy for at least one of the frequency bands is determined. A respective local signal-to-noise ratio for at least one of the frequency bands is determined as a function of the respective signal energy and the respective current estimate of the noise energy. A respective smoothed signal-to-noise ratio is determined from the respective local signal-to-noise ratio and another respective signal-to-noise ratio estimated for a previous frame. A respective filter gain value is calculated for the frequency band from the respective smoothed signal-to-noise ratio. Also, it is determined whether at least a respective one as a plurality of frames is a non-speech frame. When the frame is a non-speech frame, a noise energy level of at least one of the frequency bands of the frame is estimated. The band is filtered as a function of the estimated noise energy level.

Case	Conditions to be met	value of update constant α
Watch dog timer expired	$UpdateTimer > TimeOut$ $PredErr > T_{PE1}$	$\alpha = 0.002$
Frame is stationary	Stationary counter is > 0.5 sec.	$\alpha = 0.05$
Frame speech likelihood indicates non-speech	$NoiseLikelihood < T_{LTK}$ $PredErr > T_{PE2}$	$\alpha = 0.1 \cdot NoiseLikelihood$ with $\alpha \leq 0.1$
The LPC residual of the frame has near-zero skewness	$ \gamma_3 < T_a, \gamma_3 < T_b,$ $PredErr > T_{PE2}$	$\alpha = 0.05$
Noise energy is dropping	Current noise energy estimate > total energy	$\alpha = 0.1$

TABLE 1

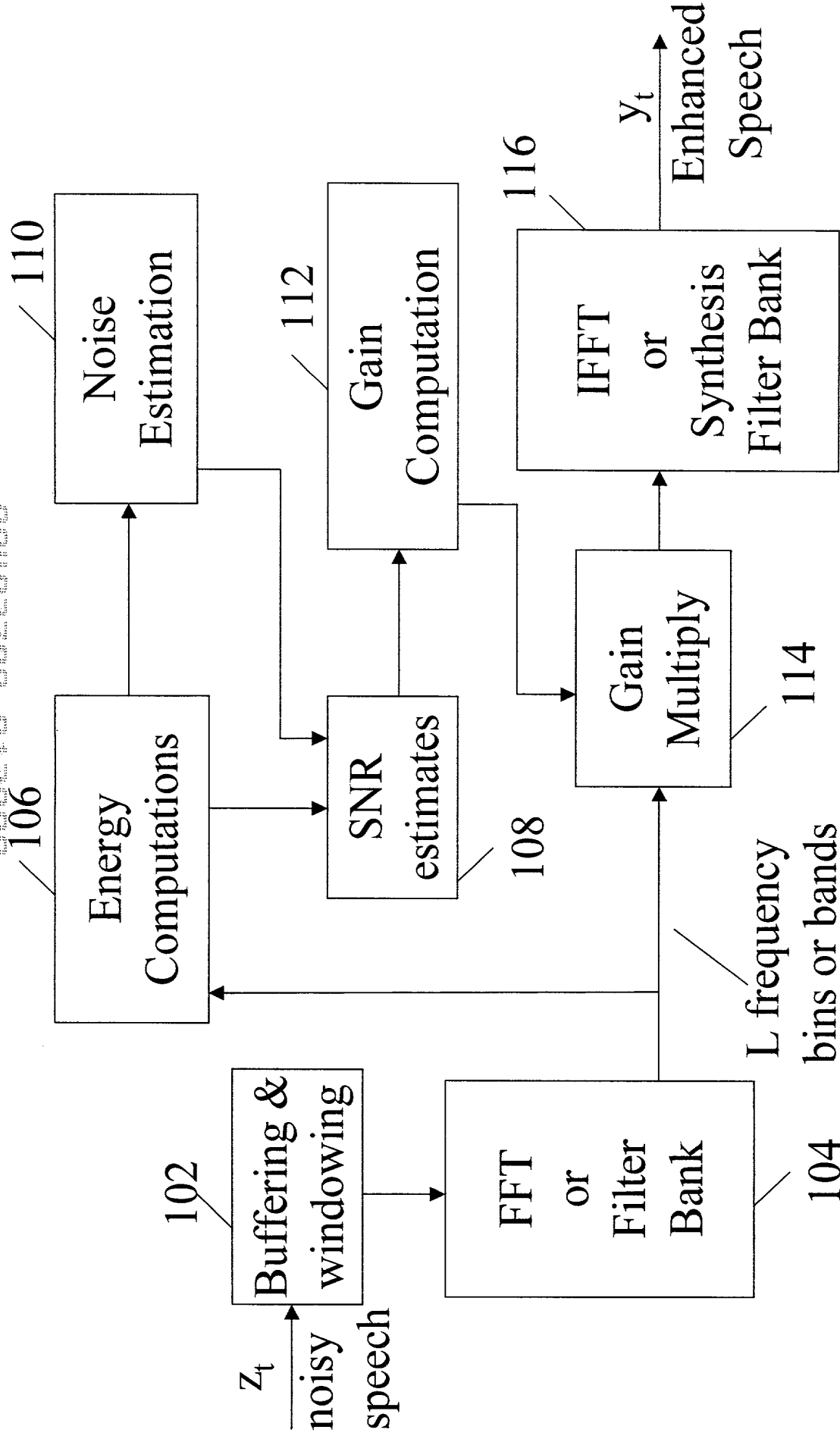


FIG. 1

RELATED ART

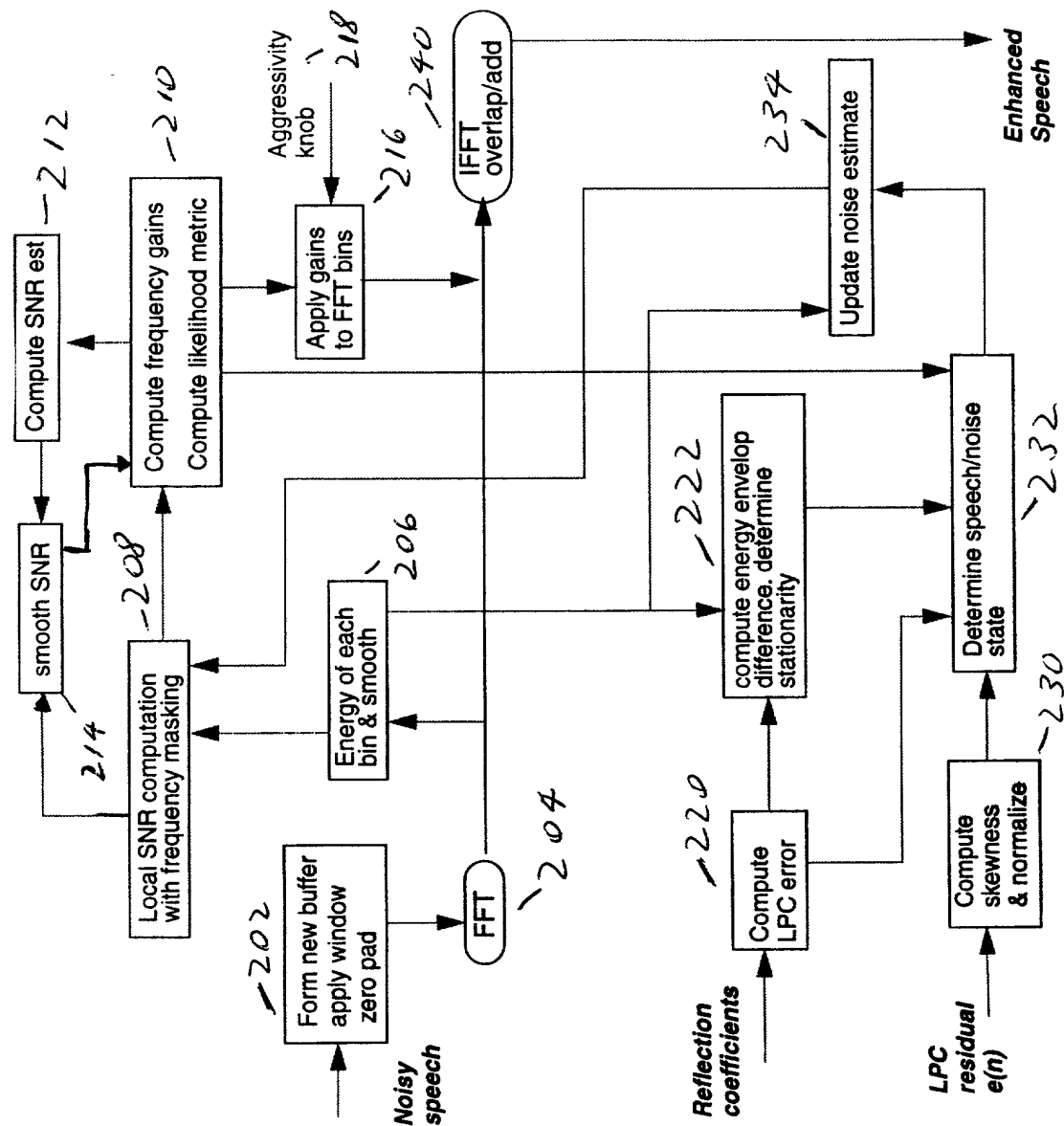


FIG. 2

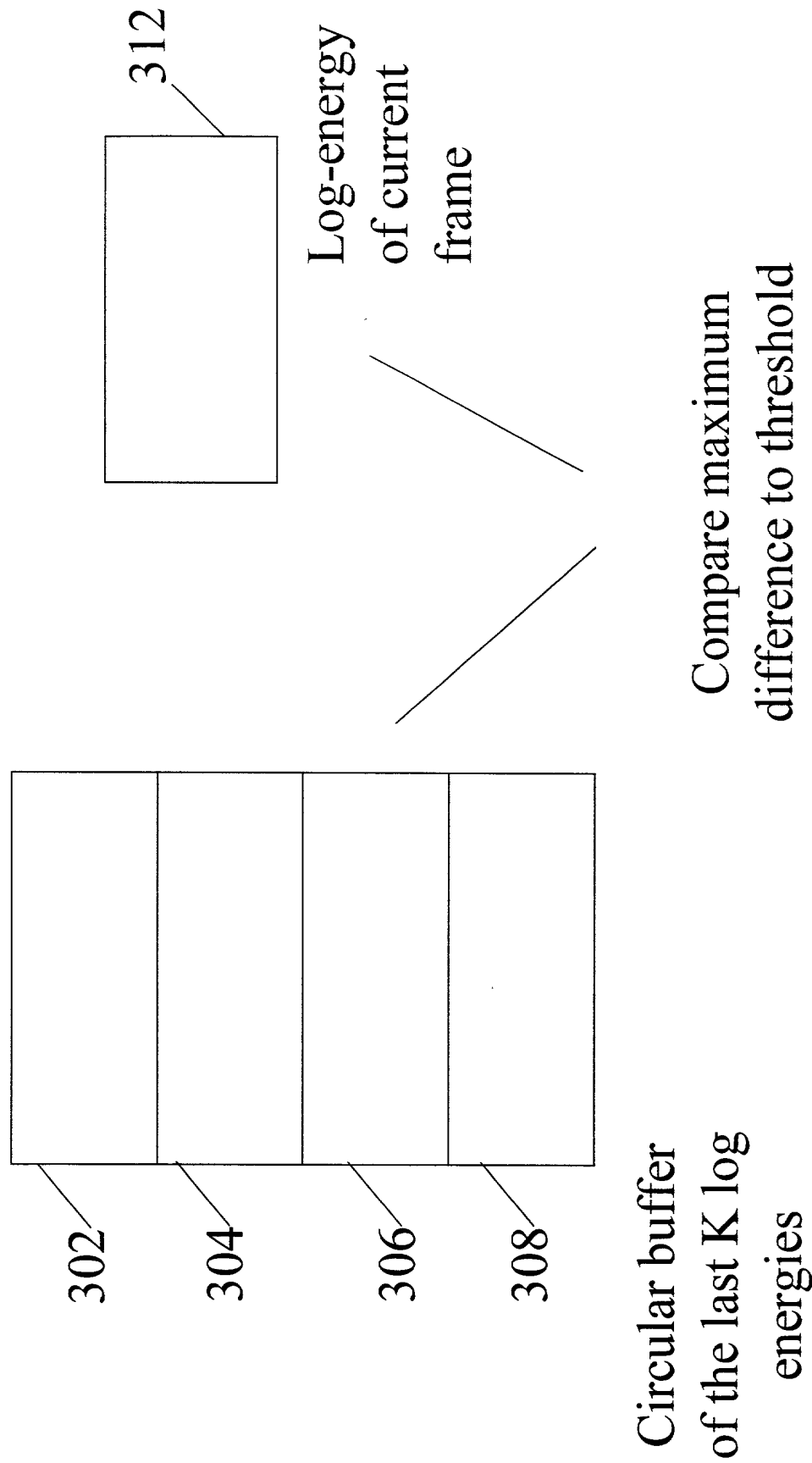


FIG. 3

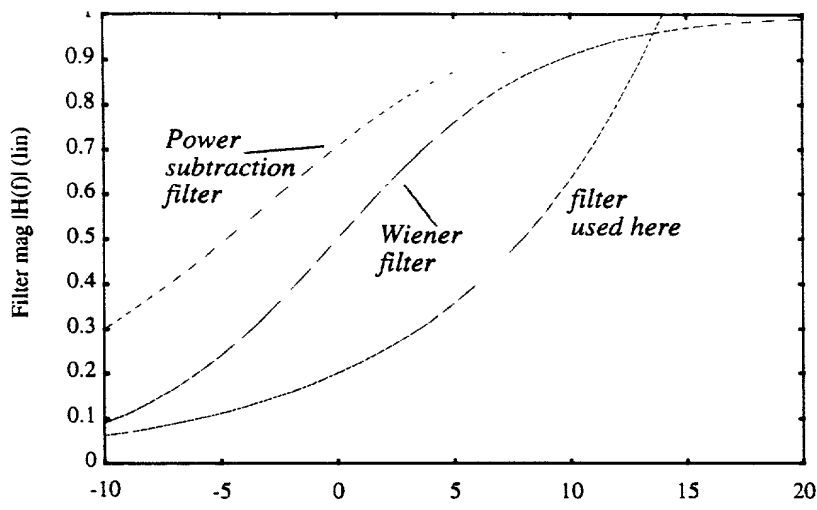


FIG. 4A
dB scale

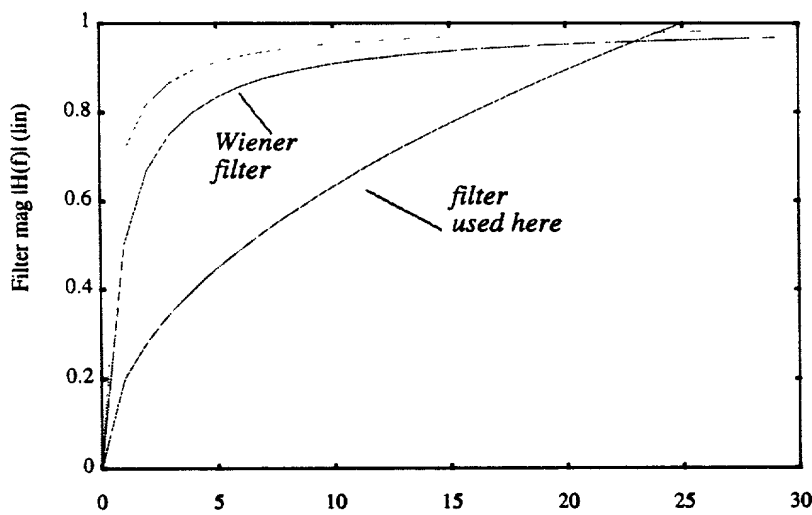


FIG. 4B
Linear scale

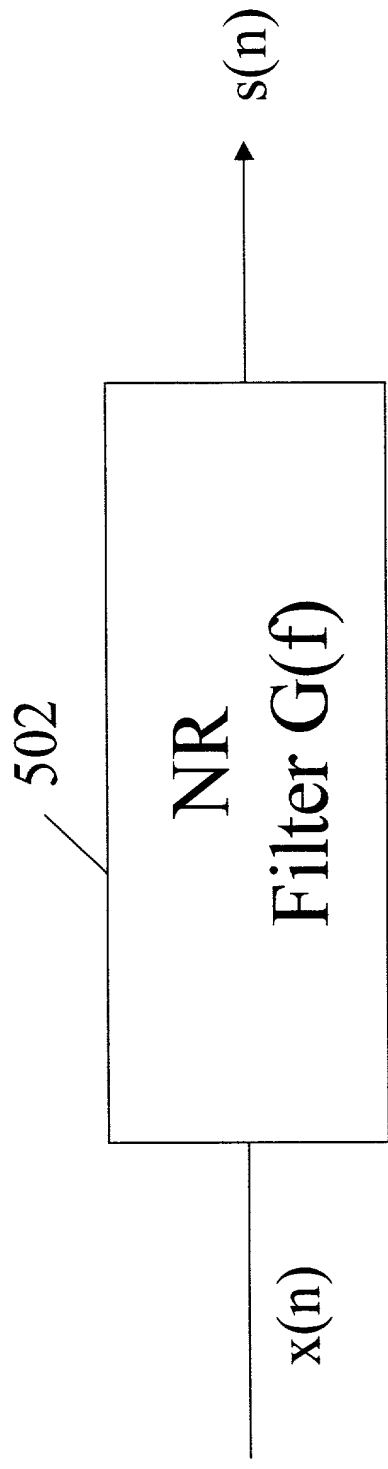


FIG. 5

DECLARATION FOR PATENT APPLICATION

As a below named inventor, I hereby declare that.

My residence, post office address and citizenship are as stated below next to my name

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

METHOD OF AND APPARATUS FOR REDUCING ACOUSTIC NOISE IN WIRELESS AND LANDLINE BASED TELEPHONY
specification of which is attached hereto unless the following is checked

[] was filed on _____ as United States Application Number or PCT International Application Number _____
and was amended on _____ (if applicable)

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above. I acknowledge the duty to disclose information which is material to patentability as defined in Title 37, Code of Federal Regulations, § 1.56.

I hereby claim foreign priority benefits under Title 35, United States Code, § 119(a)-(d) of any foreign application(s) for patent or having a filing date before that of the application on which priority is claimed

Prior Foreign Application(s)

Priority Claimed

_____ (Number)	_____ (Country)	_____ (Day/Month/Year Filed)	<input type="checkbox"/> Yes <input type="checkbox"/> No
_____ (Number)	_____ (Country)	_____ (Day/Month/Year Filed)	<input type="checkbox"/> Yes <input type="checkbox"/> No

I hereby claim the benefit under Title 35, United States Code, § 119(e) of any United States provisional application(s) listed below

_____ (Application Number)	_____ (Filing Date)
_____ (Application Number)	_____ (Filing Date)

I hereby claim the benefit under Title 35, United States Code, § 120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, § 112, I acknowledge the duty to disclose information which is material to patentability as defined in Title 37, Code of Federal Regulations, § 1.56 which became available between the filing date of the prior application and the national or PCT International filing date of this application

_____ (Application Number)	_____ (Filing Date)	_____ (Status - patented, pending, abandoned)
_____ (Application Number)	_____ (Filing Date)	_____ (Status - patented, pending, abandoned)

I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith Peter T Cobrin, Reg. No. 24,117, Marvin S Gittes, Reg. No. 24,350, Richard M. Lehrer, Reg. No. 38,536, Robert J. Hess, Reg. No. 32,139, David W. Denenberg, Reg. No. 40,986, Michael A. Adler, Reg. No. 38,810, Gerald J. Cechony, Reg. No. 31,335, Lawrence E. Russ, Reg. No. 35,342.

Address all correspondence to COBRIN & GITTES
750 Lexington Avenue, New York, New York 10022 Telephone: (212) 486-4000 Facsimile: (212) 486-4007

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true, and further that these statements were made with the knowledge that willful false statement and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon

Full name of sole or first inventor (given name, family name) Elias J. Nemer

Inventor's signature _____ Date: _____

Residence: Montreal, Quebec, Canada

Post Office Address: 3475 De La Montagne, Apt. 704, Montreal, PQ, H3G 2A4 Canada

Citizenship: Canada

Full name of second joint inventor (given name, family name) _____

Inventor's signature _____ Date: _____

Residence: _____

Post Office Address: _____

Citizenship: _____

Full name of third inventor (given name, family name) _____

Inventor's signature _____ Date: _____

Residence: _____

Post Office Address: _____

Citizenship: _____

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statement and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon

_____ Additional inventors are being named on separately numbered sheets attached hereto.